
MAUSA

Author: Philip J. Uren
Date: 15th September 2005
Version: 1.0a
Platform: Win/Dos
Files: MAUSA.EXE
BLOSUM_COST\BLOSUM50
BLOSUM_COST\BLOSUM62
BLOSUM_COST\BLOSUM80

Disclaimer

We provide this program, MAUSA, as is and with no warranty or guarantees (implied or otherwise). We accept no responsibility for any losses that may occur as a result of the use or misuse of this software.

Bugs/Comments

There are no known current bugs.
Please send comments and/or bug reports to: puren@utas.edu.au

When reporting bugs, be as specific as possible. Please include both the input that you used and the command line options specified.

Command Line

The following command line formats, where % is your prompt, are currently allowed:

% MAUSA.EXE input.fa
Runs MAUSA on the sequences contained in 'input.fa'.
% MAUSA.EXE input.fa -matrix user-matrix.dat
Runs MAUSA on the sequences contained in 'input.fa', using the cost-based substitution matrix defined in 'user-matrix.dat'.
% MAUSA.EXE -batch batch_input.dat
Runs MAUSA on the list of input files specified in the file 'batch_input.dat'.
% MAUSA.EXE -batch batch_input.dat -matrix user-matrix.dat
Runs MAUSA on the list of input files specified in the file 'batch_input.dat', using the matrix defined in 'user-matrix.dat'.

Input/Output Formats

MAUSA accepts sequence input in FASTA format. Please ensure that the alphabet used in your sequences matches that of the substitution matrix you have provided. See below for the format of user supplied matrices. Note that MAUSA is case-sensitive and will not regard letters of differing case as being identical (e.g. 'A' is distinct from 'a'). The matrices distributed with MAUSA use only uppercase letters, if you do not provide an alternate matrix, ensure all your sequences are in uppercase only. Please also ensure that all white space is removed from your sequences. Each sequence should be preceded by a name, indicated by beginning the line with a greater-than symbol, '>'. Ensure that there are no gap ('-') symbols in your input sequences.

MAUSA cannot process input files where either one or both of the following are true:

- a.) There are two or more identical sequences

b.) There are two or more identical names
The performance of the program in such a situation is undefined

Alignments will be produced in FASTA format. The sequence order will not necessarily be the same as that in the input file. All sequences will be named as they were in the input file. Note that sequence names are treated as literal values and no dynamic information (for example sequence length) within them will be updated. Output files will have the same name as input files, with ".aln" appended to the end.

Batch File Format

When running the program in batch mode, the user must supply a list of files to process. This list is contained within a plain text file - ensure that there is no unnecessary white space. The format is simply a list of filenames. For example:

```
input\1ar1.fa
input\1aboA.fa
input\1idy.fa
input\2myr.fa
```

Will process the files '1ar1.fa', '1aboA.fa', '1idy.fa', '2myr.fa' in the folder 'input'. Assuming the process completes successfully, this will produce the output files in the same directory with the same names, but '.aln' appended to each.

Substitution Matrices

MAUSA uses cost based substitution matrices. Blosun50, 62 and 80 matrices are distributed with the program in this format, but the user may also specify an alternate matrix. For an example of the required format, please view any of the enclosed matrices. Note that MAUSA provides a mechanism for selecting an appropriate matrix from the three default matrices provided and scaling the gap initiation cost. Should you supply your own matrix, this feature will be disabled. An example of the format is as follows:

```
-----
A  R  N  D
0
11 0
14 11 0
14 15 10 0

19 7
-----
```

The alphabet is 'A', 'R', 'N' and 'D'. The cost of substituting any amino acid with itself is 0. 'A' to 'R' is 11, 'R' to 'D' is 15, etc. Note that the matrix is symmetric: 'X' to 'Y' is considered equal to 'Y' to 'X'. The last two numbers are the gap initiation cost (gamma) and the gap extension cost (lambda) respectively.

Revision History

V1.0a

- Added command line switches
- Reduced status messages
- Added failsafe to stop MAUSA from attempting to force a tree modification that's larger than is possible.